

SUBJECTIVE APPRAISAL AS A FEEDBACK TOOL

Billy L. Burnside

U. S. Army Research Institute¹
Fort Knox Field Unit

The products of U. S. Army Centers/Schools are trained graduates and training support materials. In order to appraise the quality and utility of these products, training developers and evaluators in the Centers/Schools need meaningful feedback from users at the institution and in the field. There are six principle methods which these personnel may use to obtain such feedback: receipt of informal comments, administration of surveys/questionnaires, conduct of interviews, analysis of available unit performance records, observation of training classes and exercises, and administration of performance tests. Interviews with battalion commanders and staffs (Burnside, 1981) and with training developers and evaluators in a typical Center/School (Witmer and Burnside, 1982) indicate that the first three of these methods are the most commonly used. A common attribute of these three methods is that they are relatively subjective in nature; i.e., they are largely based upon individuals' perceptions, judgments, and opinions.

Since the feedback presently available to training developers and evaluators consists largely of subjective data, an important issue to be addressed is how accurate or valid these data are. That is, how do they compare with data gathered using more objective methods and criteria? This issue is addressed in the present paper by reviewing research results comparing subjective ratings gathered using surveys or interviews with relatively objective data gathered using structured observations or "hands-on" performance tests. The type of feedback of interest here is appraisal of the performance of individual soldiers and military units on specific tasks, rather than assessment of general knowledge and abilities. An example of subjective appraisal is using a survey or interview to ask a soldier whether he or she can perform a specific task. The comparable objective appraisal would involve administration of a "hands-on" test in which the soldier's performance was compared to a validated standard. Subjective appraisal is a relatively efficient and cost-effective method of gathering feedback, so it will continue to be used in the military. The key question thus becomes whether the data gathered using this approach are sufficiently accurate to warrant their use in particular situation, and whether their accuracy can be increased by refinements in collection methodologies.

The aspects of subjective feedback addressed in this paper include what is appraised, who does the appraising, and how the appraisal is done. The type of appraisal of greatest interest here involves estimates of soldiers' proficiencies on specific tasks, but other types addressed include judgments of the criticality, difficulty, and performance frequency of specific tasks.

¹The views expressed in this paper are those of the author and do not necessarily reflect the view of the U. S. Army Research Institute or the Department of the Army.

These are the types of estimates typically obtained using Comprehensive Occupational Data Analysis Program (CODAP) surveys. The issue of who does the appraisal is addressed by summarizing research relating to self-appraisals, supervisory appraisals, and peer appraisals. Discussion of the issue of how subjective appraisals are done centers around survey and interview techniques, and the paper concludes with discussion of ways to improve the accuracy of subjective data.

Types of Appraisals

Proficiency

A key element of feedback to Army Centers/Schools is data relating to the proficiency with which soldiers can perform specific required tasks. Such data are needed to allow training developers to evaluate both institutional and unit training and to make modifications as needed. Since the operational testing of soldiers' performance is costly in terms of time and resources, proficiency data are usually gathered through subjective estimates. That is, soldiers are asked to estimate their confidence or the likelihood that they can perform specific tasks. Supervisors may also be asked to rate soldiers' proficiencies. How accurately do such subjective appraisals reflect actual task proficiencies?

Several pieces of research conducted outside the military are relevant to answering this question. There is some evidence that people can appraise their own task-specific proficiencies with moderate accuracy, as long as the tasks appraised are basic ones with which they have had extensive experience. For example, Ash (1980) found that self-ratings of straight copy typing ability correlated in the .44 to .59 range with the results of typing tests. However, subjective ratings of more complex typing skills did not correlate as highly with performance. In a recent meta-analysis of self-evaluation of ability, Mabe and West (1982) found the mean correlation between self-evaluation and performance measures to be approximately .30. While they found many methodological weaknesses that limited the interpretation of correlational data, the general conclusion is that self-appraisals of proficiency are not particularly accurate. In a meta-analysis of educational research, Cohen (1981) found that the mean correlations between students' subjective appraisals of instruction and measures of students' proficiencies ranged from .38 to .47. He also identified several methodological problems, such as the lack of objective criteria to compare subjective appraisals against and the fact that most appraisals obtained have been global rather than task-specific in nature. DeNisi and Shaw (1977) avoided some of the common methodological problems by examining the accuracy of self-appraisals for specific abilities on tasks such as visual pursuit and manual speed and accuracy. While the correlations between self-appraised and tested abilities were almost all statistically significant (in the .20 to .40 range), they showed that these results had little practical significance. Due to methodological weaknesses in the relevant research and problems in interpreting correlations in the .30 to .40 range, the appropriate conclusion appears to be that there is no convincing evidence that subjective appraisals of proficiency are accurate.

Few studies of the accuracy of subjective appraisals of proficiency have been conducted in a military setting. Many of those that have been conducted

have suffered from methodological problems, such as the lack of objective criteria or the lack of specificity or explicitness in the tasks addressed. For example, Hall, Denton, and Zajkowski (1978) found that supervisors' estimates of sailors' proficiencies on several tasks did not correlate significantly with performance. However, the criterion used was performance on a written test rather than "hands-on" performance. A further examination of two sets of data previously published by the Army Research Institute provides some insights that have not previously been available.

Hiller (1980) collected data which allow comparison of self-estimates and "hands-on" performance test results for five specific tasks. The general finding is that self-appraisals of proficiency were accurate for general leadership skills, were at best moderately accurate for cognitive skills, and were inaccurate for motor skills. The accuracy of subjective appraisals was thus found to decline as the objectivity of the performance test criterion and standards increased. Leadership skills are difficult to develop standards for and objectively evaluate; the high accuracy of self-appraisals of leadership skills may have resulted from the comparison of these appraisals with results of relatively subjective performance tests. Relatively objective performance tests are available for "hands-on" motor skills, and self-appraisals of such skills were highly inaccurate. This indicates that subjective appraisals of proficiency are not accurate when compared to an objective criterion.

In the military skill retention literature, several instances can be found in which self-appraisals of proficiency were collected prior to a retention test, but the results were not reported. This leads one to suspect that the results were negative; i.e., that the self-appraisals were not found to be accurate. This suspicion is supported by further examination of data collected by Shields, Goldberg, and Dressel (1979), in which confidence ratings of proficiency on 20 tasks were found not to significantly correlate with performance test results. It thus appears that retention research has not supported the accuracy of subjective appraisals of proficiency.

The data reviewed above indicate that subjective appraisals of proficiencies (largely in terms of self-appraisals) on specific tasks often do not represent true abilities. This appears to be especially true when the subjective appraisals are compared to objective well-specified performance criteria. Before subjective appraisals are used as feedback to training developers, the relationship between such appraisals and more objective measures of performance should be further examined. Self-ratings of proficiency may only be accurate when addressing explicit tasks with which the ratees have extensive experience.

Criticality

Since training resources are limited, training developers must somehow determine which tasks are most critical for combat performance and therefore most important to train. This is typically accomplished by preparing an extensive list of tasks and asking subject matter experts to subjectively rate their criticality. Just as with estimates of proficiency, one can question how accurately subjective appraisals of criticality represent the "true" relative importance of tasks. Data are relatively sparse in this area, but those available indicate that rater agreement (interrater reliability) has generally been found to be low. The accuracy or predictive validity thus would be

expected to be low. Another problem in this area is the specification of an objective criterion of criticality. Due to these reliability and criterion problems, subjective appraisals of task criticality should be used cautiously, if at all.

Difficulty

Knowledge of the relative difficulty of tasks is important to training developers, in order to determine the proper distribution of training time and resources. Appraisals of task difficulty are usually made subjectively, based upon the experiences and opinions of subject matter experts. Indications are that subjective appraisals of task difficulty are not generally accurate; i.e., the tasks picked as most difficult by subject matter experts are not the ones most commonly failed by soldiers. Part of the reason for this problem may lie in the fact that difficulty is not consistently defined. Some tasks are difficult to learn but not to perform, and vice versa. Raters having different perceptions of what is meant by difficulty would thus provide unreliable ratings for such tasks. In obtaining subjective appraisals of task difficulty, care must be taken to precisely define the rating dimension.

Frequency

While limited relevant data are available, indications are that judgments of the frequency with which specific tasks are performed are not generally accurate. Again, there is a criterion problem here, since objective measures of task performance frequency can only be obtained through laborious observation in the field. In cases where this has been done (e.g., Johnson, Tokunaga, and Hiller, 1980), accurate frequency estimates have been obtained only for broad categories of tasks addressed through carefully controlled data collection techniques. As with the other types of subjective appraisal addressed above, frequency estimates should not be assumed to be accurate. They should be collected very carefully and their accuracy should be checked against objective criteria.

Types of Appraisers

A primary consideration in the use of subjective appraisals is the sources from which they are collected. Three general alternative sources are available for providing subjective appraisals as feedback: soldiers evaluating themselves (self-appraisal), supervisors, and peers. Research on the relative accuracy of these appraisal sources has produced mixed results; it is difficult to address the relative accuracy of these sources when the absolute accuracy of each of them is undetermined.

The biggest advantage of self-appraisals is that individuals have extensive data available about themselves and can provide information that is unavailable from other sources. Individuals are aware of situational factors in their own behavior, and are less likely to over-generalize than outside observers are. A problem with self-appraisals is that individuals may not be capable of appraising themselves accurately, as shown by the research summarized in the previous section. Another problem is that individuals may have reason to bias their self-appraisals in a positive direction, resulting in leniency errors. Such errors are common in self-appraisals, but they can be reduced by techniques such as making the appraisals publicly verifiable

(van Rijn, 1981). When self-appraisals are used, their accuracy should be checked against an objective criterion, and the appraisers should be aware that this is being done.

The research literature does not at this time allow any definitive conclusions on the relative accuracy of subjective appraisal sources. What is needed is a study which includes the collection of supervisory, peer, and self-predictions of proficiencies on specific tasks, followed by objective measures of task performance. The literature thus far has generally failed to include objective criteria for comparison purposes, and until it does the accuracy issue will be unresolved. Self-appraisals often suffer from leniency biases, and peer and supervisory appraisals may suffer from tendencies to over-generalize from small samples of data. Accuracy of these approaches should thus not be assumed, but should be checked against relatively objective criteria.

Types of Appraisal Methods

The previous discussion leads to two primary conclusions about subjective appraisal. The first of these is that adequate data are not yet available to determine either the absolute accuracy of subjective appraisals or the relative accuracy of different appraisal sources. The second is that the limited research which has directly addressed the accuracy of subjective appraisals has in general not found it to be high. These appraisals should thus be used very cautiously with frequent checks on their accuracy. However, military agencies will continue to use subjective appraisals as feedback, due to the ease with which they can be collected. Recognition of this fact leads to the need to identify ways in which the accuracy of subjective appraisals can be increased. A review of the literature by the present author and a meta-analysis reported by Mabe and West (1982) has indicated several ways in which this can be done. These are briefly summarized below.

1. Integrate mutually supportive subjective appraisal methods within a feedback system. Since no appraisal method is complete and sufficient in and of itself, methods should be used to complement each other. Surveys can be used to obtain a general overview of the situation, interviews can be used to obtain more in-depth detail on specific problems, and observations and performance tests can be used as accuracy checks.
2. Ensure that question developers and subjective appraisers have a common basis of understanding. These groups should share a common understanding of task elements, successful task completion, appropriate standards, and rating dimensions.
3. Design questions to maximize accuracy. Make the situation and behavior being addressed as explicit as possible, and specifically state the action being addressed.
4. Make rating scales as explicit as possible. Phrase rating scales in terms of observable measures of performance, rather than in vague, general terms.
5. Be sure that raters have had experience with the tasks rated. Ensure that supervisors have had ample opportunity to observe task performance by the people they are rating.

6. Train raters before they provide subjective appraisals. This training should include experience with the rating scales to be used, a discussion of common types of psychometric errors, and a discussion of the dimensions of the situation being evaluated.

7. Facilitate raters' recall of relevant experiences. Ask raters to review their previous experiences, provide them with thorough descriptions of the tasks and situations being rated, and provide any other cues which aid memory.

8. Make certain that appraisers have the cognitive capacity and motivation to provide accurate ratings. Explain the need for accurate rating data during instructions. Check the accuracy of subjective ratings whenever possible, and let the raters know that this will be done.

References

- Ash, R. A. Self-assessments of five types of typing ability. Personnel Psychology, 1980, 33, 273-282.
- Burnside, B. L. Field performance feedback - A problem review (Draft Research Report 1323). Alexandria, VA: US Army Research Institute, August 1981.
- Cohen, P. A. Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. Review of Educational Research, 1981, 51, 281-309.
- DeNisi, A. S. & Shaw, J. B. Investigation of the uses of self-reports of abilities. Journal of Applied Psychology, 1977, 62, 641-644.
- Hall, E. R., Denton, C. F., & Zajkowski, M. M. A comparative assessment of three methods of collecting training feedback information (TAEG Report No. 64). Orlando, FL: Training Analysis and Evaluation Group, December 1978.
- Hiller, J. H. A methodology for estimating the cost-effectiveness of alternative pretesting procedures (Draft Technical Report). Alexandria, VA: US Army Research Institute, November 1980.
- Johnson, C. A., Tokunaga, H. T., & Hiller, J. H. Validation of a job analysis questionnaire against intensive observation. Paper presented at the Military Testing Association Conference, Toronto, October 1980.
- Mabe, P. A. III & West, S. G. Validity of self-evaluation of ability: A review and meta-analysis. Journal of Applied Psychology, 1982, 67, 280-296.
- Shields, J. L., Goldberg, S. L., & Dressel, J. D. Retention of basic soldiering skills (Research Report 1225). Alexandria, VA: US Army Research Institute, September 1979.
- van Rijn, P. Self-assessment in personnel selection and placement. Paper presented at the Military Testing Association Conference, Arlington, VA, October 1981.
- Witmer, B. G. & Burnside, B. L. Feedback needs of training developers and evaluators (Draft Research Report). Alexandria, VA: US Army Research Institute, September 1982.